Inventory and Monitoring Program

National Park Service
U.S. Department of the Interior

Data Management Plan

for the

Cumberland Piedmont Network

and

Mammoth Cave National Park Prototype Monitoring Program

Compiled by:

Bill Moore
Mammoth Cave National Park
Mammoth Cave, Kentucky 42259

Rob Byrd
Department of Computer Science
Western Kentucky University
Bowling Green, Kentucky
42101

Teresa Leibfreid
Cumberland Piedmont Network
Mammoth Cave, Kentucky 42259

September 30, 2005

Acknowledgements

This data management plan benefited greatly from the collective efforts of the Inventory and Monitoring Program's – Data Management Planning Work Group. This group of data managers from within the I&M Program held numerous conference calls, submitted draft materials, and provided guidance and suggestions to the 12 "first year" networks developing their data management plans. Gary Enstminger provided valuable editorial assistance to the work group. In addition, we are also indebted to Steve Fancy, Joe Gregson, and Lisa Nelson for their willingness to review and provide feedback on draft materials. We also extend our gratitude to Pat Price and Matt Arnold, Mammoth Cave National Park Information Technology staff, for their guidance and assistance in this endeavor as well as the editorial assistance and feedback provided by Steve Thomas and Douglas Foster. Dr. Rob Byrd's assistance in this endeavor was made possible through the Southern Appalachian Cooperative Ecosystem Studies Unit. Thank you, one and all.

Table of Contents

List of Tables

List of Figures

Executive Summary

As part of the Natural Resource Challenge, the National Park Service (NPS or Service) has implemented a strategy to institutionalize inventory and monitoring across the Service. National level program coordination and management for this strategy is being provided by the NPS Inventory and Monitoring Program (I&M Program). The strategy consists of a framework which includes the establishment of experimental prototype monitoring programs and grouping of parks into networks based on geography and similarities in natural resource characteristics. The Cumberland Piedmont Network (CUPN) and Mammoth Cave National Park (MACA) Prototype Monitoring (Prototype) Program, collectively referred to as CUPN-MACA, was established to support the long-term inventory and monitoring goals of the I&M Program.

Data Management Rationale

Data management is/will be addressed at three levels of detail by CUPN-MACA (Figure 1). Each network is required to complete a Network Vital Signs Monitoring Plan. Networks must receive approval on that plan from the national monitoring program leader before implementation can commence. Chapter VI (Data Management) of the Network Vital Signs Monitoring Plan is intended to provide summary information excerpted from the respective network's DMP. In addition, the DMP will be attached as an appendix to the Network Vital Signs Monitoring Plan. This approach ensures networks prioritize data management planning early on in program development.

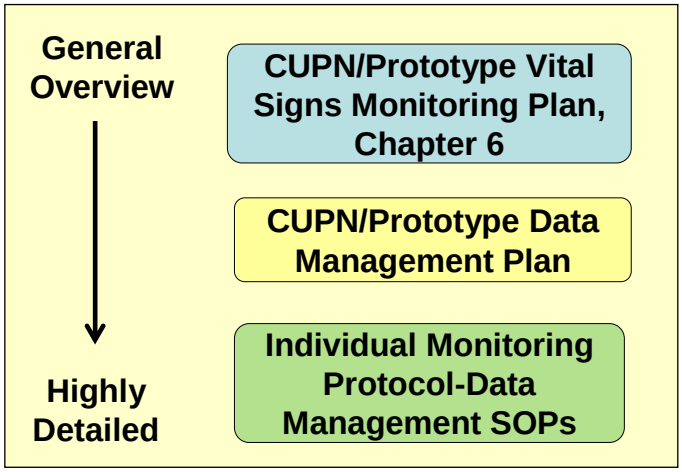

Figure 1. Data Management Guidance for CUPN-MACA

In 2004, 12 "first year" networks, including CUPN-MACA, collaborated in the development of draft data management plans (DMPs) for their respective networks, with initial coordination and guidance being provided by the I&M Program. While data management planning is in many ways a new endeavor to the NPS, its importance cannot be overemphasized. Park managers and policy makers require timely, credible information in a useable format if they are to fulfill the NPS mission. The primary purpose of the CUPN-MACA DMP is to communicate an overarching data management strategy that establishes guidance and specific policy, as appropriate. Specific goals include:

- Develop a data management process that supports and enhances the inventory and long-term ecological monitoring goals and objectives of the Network, Prototype, and I&M Program.
- Ensure adequate hardware and software resources (tools) for managing data are available.
- Maintain properly trained staff members that understand their roles and responsibilities of data collection, entry, analysis, and reporting.

- Ensure the long-term integrity and availability of data products produced and/or utilized by CUPN-MACA.
- Facilitate the adoption and use of high quality data management principles, policies, and procedures as an integral part of day to day CUPN-MACA activities.

Data Management Structure/Design

Managing data is the shared responsibility of everyone involved with data, from producers to end-users. Effective data stewardship is dependant upon an effective organization with well defined roles and commensurate responsibilities. Within CUPN-MACA's framework of roles and responsibilities are particular "core roles" for effective project-level data management. These include the project leader, who oversees and directs day-to-day project operations; the data manager, who ensures data are organized, useful, compliant, safe, and available; and the GIS specialist, who incorporates and manages spatial data. Each is responsible for certain aspects of project data, and all share responsibility for some shared tasks.

Another important framework for effective data management is the computers and servers linked through computer networking services. The I&M Program data management framework is the conduit through which most, if not all, CUPN-MACA data will flow (Figure 2). The framework includes a series of internet-based, master databases to promote integration and enable linkages and data sharing to other external databases including NPS permitting system, Integrated Taxonomic Information System (ITIS), USFWS T&E species database, NatureServe, and eNature. A second component of the data management framework is a series of desktop applications in MS Access (the NPS standard for desktop relational databases) that can accommodate the same data as the master web-based databases. The desktop versions increase the availability and utility of each park's data by allowing users (with the appropriate permissions) to download the latest data from the corresponding master web database and develop customized displays or analysis of the data or integrate them with other local datasets. A third component of the framework is a collection of relational databases that follow the Natural Resource Database Template scheme, with an integrated link to GIS and associated tools through an Arc-Access Link Tool or geodatabase model.

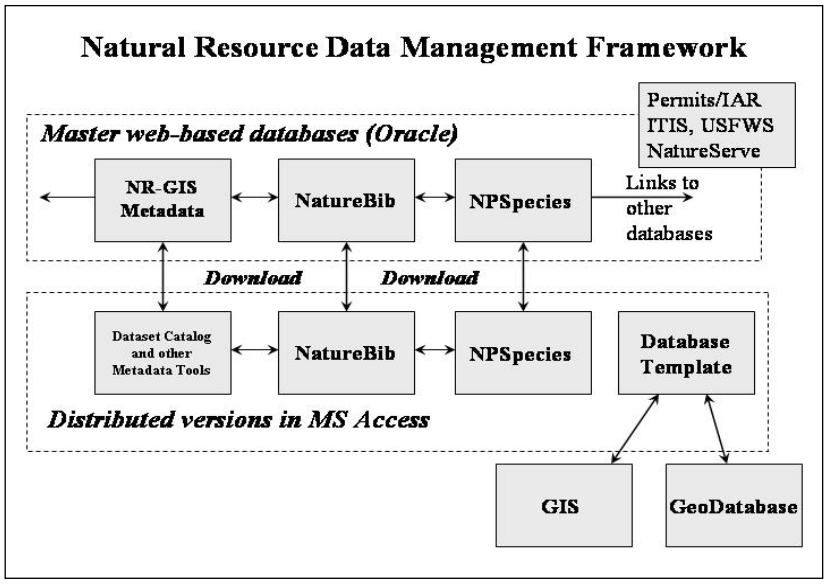

Figure 2. National I&M Program Data Management Framework

Rather than developing a single, integrated database system, our data design relies upon modular, standalone project databases that share design standards based on the Natural Resource Database Template and links to centralized data tables. Each project database contains three primary components.

- Common lookup tables – Links to entire tables that reside in a centralized database, rather than storing redundant information in each database. These tables typically contain information that is not project-specific (e.g., lists of parks, personnel, and species).
- Core tables and fields based on CUPN-MACA and national templates – These tables and fields are used to manage the information describing the "who, where and when" of project data. Core tables are distinguished from common lookup tables in that they reside in each individual project database and are populated locally. These core tables contain critical data fields that are standardized with regard to data types, field names, and domain ranges.
- Project-specific tables and fields – The remainder of database objects can be considered project-specific, although there will typically be a large amount of overlap among projects. This is true even among projects that may not seem logical – for example, a temperature field will require similar data types and domain values. As much as is possible, efforts will be made to develop these project-specific objects to be compatible with those maintained by other networks and cooperators managing similar datasets – particularly if integration with other planned or existing data sets is important for meeting project objectives.

Data Management Implementation

CUPN-MACA projects can be divided into five primary stages:

- planning and approval
- design and testing
- implementation
- product integration
- evaluation and closure

Each stage is characterized by a set of activities carried out by staff involved in the project. Primary responsibility for these activities rests with different individuals according to the different phases of a project. It should be noted, the quality, integrity, and usability of the resulting data depend significantly on whether the project leader and data manager devote adequate attention to properly developing the sampling plan (including sampling parameters), and the project-specific tables and fields during the design and testing phase. Devoting adequate attention to this aspect of the project is possibly the single most important part of assuring the quality, integrity, and usability of the resulting data.

Data Acquisition

There are many potential sources of important data and information about the condition of natural resources in CUPN parks. In a broad sense, these sources can be divided into three broad categories:

- Program data - Data generated or funded by CUPN-MACA in support of I&M Program goals. Primary datasets include the 12 basic resource inventories conducted by the I&M Program and CUPN-MACA vital signs monitoring protocols.
- Non-program, NPS Data - Data generated by personnel involved in projects initiated at the individual park level or by other NPS regional or national programs (i.e., Air Resources Division, Exotic Plant Management Teams, etc.)
- Non-program, External Data – Data generated by entities external to NPS. Such data *may* not be directly linked to CUPN parks but may instead pertain to methodologies or protocols that assist in providing a regional context to park natural resource condition, threats, and trend.

Our challenge is to identify, prioritize, and acquire useful datasets; and transform them into useable formats. CUPN-MACA data acquisition and processing efforts can be broken down into three broad steps:

1. Identify, generate and/or collect data from multiple sources.
2. Ensure data are in compliance with program standards and formats.
3. Incorporate data into program data holdings as appropriate.

If data are to be transformed into useful information, the user must know something about the dataset's quality, as well as context (e.g., who collected it, when, for what purposes). Thus acquisition and processing should not be viewed as a stand-alone process - completely separate from data quality and documentation.

Data Quality

The overarching goal in establishing data quality is to ensure that a project produces data of the right type, quality, and quantity to meet project objectives and the user's needs. CUPN-MACA believes the most effective mechanism for ensuring these parameters are achieved, is to provide procedures and guidelines to assist individuals in accurate data collection, entry, and validation. Therefore, a comprehensive set of SOPs and/or other project specific guidance will be written and include clear field methodologies, staff training, well-organized field forms, and data entry applications with simple built-in validation.

The data entry programs, written in MS Access, are designed with both data quality and data security in mind. Only data managers have permission to add personnel to the data entry list. Furthermore, the programs have multiple validation checks to prevent the entry of erroneous data.

Where possible, fields are automatically entered by the computer. For example, the Event ID for most protocols will be an automatically generated globally unique identifier (GUID) that is entered by the computer whenever a new "event record" is created. This ensures that the record will always contain a unique key, thus preventing possible query errors at a later time.

Where a sample characteristic datum spans a normal range, the database program checks the entered value with the minimum and maximum value for that characteristic. If an entered value is out of range, a warning message appears and asks the user to recheck the value.

In some cases, entry may be confusing, such as, when there are up to 100 observation records per plot or site, multiple observation sites per landmark, multiple landmarks per location, and multiple locations per event. The data entry programs guide the user to the proper entry record by automatically inserting new records, filling in the subsequent landmark or plot, and placing the cursor at the proper field for entering the next datum. While it takes longer to write a program in this manner, it is, in the long run, more cost effective than having to repeatedly perform 100 percent datasheet checks looking for entry errors. The confusion is further reduced by developing computer forms that mimic the field datasheets as close as possible. This also reduces eyestrain for the technicians as they enter and visually recheck the data as detailed in the data management SOPs.

Data Documentation

Data documentation is a critical step toward ensuring that datasets are useable for their intended purposes well into the future. This involves the development of metadata, which can be defined as information about the content, quality, condition and other characteristics of data. Additionally, metadata provide the means to catalog datasets, within intranet and Internet systems, thus making the respective datasets available to a broad range of potential data users.

Because metadata documentation can be accomplished in a variety of formats and levels of detail, it can become a consuming, some might say exhaustive task. As of 2004, CUPN-MACA's intended approach was to develop a simple Dataset Catalog record for relevant geospatial and non-geospatial data. This approach would provide brief metadata for all CUPN-MACA data holdings in a searchable, centralized location. However, in 2005, several milestones in implementation of the NPS Integrated Metadata System Plan were achieved, resulting in a centralized repository of metadata records from the Natural Resources and NPS GIS Programs. As a result, CUPN-MACA is currently evaluating the most efficient process for creation and long-term management of metadata. Irregardless of method, all GIS layers will be documented with applicable Federal Geographic Data Committee (FGDC) and NPS metadata standards.

Data Analysis and Distribution

There will be two main categories of data analysis conducted by CUPN-MACA. The first and only analysis available during startup years (1-5) will be annual summaries. The second type of analysis will be used to detect long-term trends and will become available after multiple years (5-10) of monitoring have been completed. Obvious exceptions will be in those cases where long-term datasets already exist, thus allowing trend analyses to be conducted sooner. Within these two categories, CUPN-MACA will develop a broad suite of reporting formats including: external comparison reports (vital sign comparisons with other local/regional studies), annual reports (park-specific summaries of vital signs monitoring), long-term trend (five to 10 year summaries focused on a specific vital sign), annual administrative reports and work plans (accomplishments and scheduled activities), and park/regional newsletter articles. In addition,

CUPN-MACA will conduct bi-annual symposia, develop brochures, and utilize websites to report on its accomplishments. In all instances, reporting formats and contents will be tailored for the intended audience(s).

One of the stated goals of the I&M Program is to "integrate natural resource inventory and monitoring information into National Park Service planning, management, and decision making." To accomplish this goal, procedures must be developed to ensure that relevant natural resource data collected by NPS staff, cooperators, researchers and the public are entered, quality-checked, analyzed, documented, cataloged, archived, and made available for management decision-making, research, and education. Providing well-documented data in a timely manner to park managers is especially important to the success of the program. CUPN-MACA will strive to ensure:

- Data are easily discoverable and obtainable.
- Data that have not yet been subjected to full quality control will not be released, unless necessary in response to a Freedom of Information Act (FOIA) request.
- Distributed data are accompanied with appropriate documentation that clearly establishes the data as a product of the NPS I&M Program.
- Sensitive data are identified and protected from unauthorized access and inappropriate use.
- A complete record of data distribution/dissemination is maintained.

CUPN-MACA will regularly provide updated information about inventories and monitoring projects, including annual reports and detailed project reports through the CUPN-MACA web site. Information on species in the National Parks, including records generated through the I&M Program, will be maintained and accessible through the NPSpecies database. Bibliographic references that refer to National Park System natural resources will be accessible through the NatureBib database. Documents, maps, and datasets containing resource information and their associated metadata, will be accessible through the Biodiversity Data Store and/or NR-GIS Data Store. Each of these databases/repositories will be available via both a secure server and a public server. The public can access all information in these databases except those records marked as "sensitive."

Data Storage and Backup

In addition to making data and information products available to current users, it is important that data remain available and secure for future uses. Until recently, CUPN-MACA utilized Mammoth Cave National Park's data server for access, storage, and archival of digital files and relied upon MACA IT system administrators for backup and security. However, at the recommendation of MACA IT system administrators, CUPN-MACA will migrate to its own server and individualized backup strategy. At the time of this writing (September 2005), a PowerEdge Server with a RAID 5 (Random Array of Independent Disks) hard drive configuration and PowerEdge 2850 server with a RAID 5 hard drive configuration and 100T internal tape backup unit is on order. This server (i.e., the CUPN-MACA server) will be integrated into the MACA Local Area Network with security and maintenance oversight

provided by MACA IT systems administrators. Full and incremental backups will be performed on all data stored on this server per an established schedule.

The server will accommodate a hierarchical, object-oriented directory structure for securely storing digital files. Master project databases, common lookup tables, program level administrative tools, final versions of project deliverables, and most non-GIS working files are all incorporated within this structure. Note: due to their resource rich requirements most "working" GIS data are maintained off the MACA server under the supervision of the GIS specialists. Some key aspects of the file management strategy include:

- Accessibility and user privileges within the CUPN-MACA parent directory are closely managed.
- Working files are kept separate from finished products.
- Finished products are accessible but maintained as read-only.
- Although conventions may be less stringent in some areas, in general, standards such as naming conventions and file structures are enforced within the parent directory.

All paper documents and specimens managed or produced by CUPN-MACA will be managed according to curatorial recommendations. For all materials submitted to the MACA Curatorial Storage Facility, CUPN-MACA will provide essential cataloging information such as the scope of content, project purpose, and range of years, to facilitate ANCS+ record creation and accession. CUPN-MACA will also ensure that materials are presented using archival-quality materials (e.g., acid-free paper and folders, polypropylene or polyethylene slide pages and photo sleeves, etc.). Specimens collected under the auspices of CUPN-MACA will be provided to the network park in which they were collected for curation, or to a repository approved by the park (where the specimens are considered on loan).